604059

COPY 7 of 7 COPIES

1

604059

[redacted]

ON THE POWER FUNCTION OF A SIGN TEST FORMED BY USING SUBSAMPLES

John E. Walsh
Douglas Aircraft Co., Inc.

P-7
1-7-4

ON THE POWER FUNCTION OF A SIGN TEST FORMED BY USING SUBSAMPLES

By John E. Walsh

Douglas Aircraft Co., Inc.

1. Summary. The sign test can be used to obtain significance tests for the population median under extremely general conditions. One disadvantage of the sign test for the median is the limited number of suitable significance levels available for a given ~~number~~ of observations. If the observations are drawn as several subsets of specified sizes, however, a variation of the sign test can be applied using order statistics of order statistics of these subsets. This test furnishes a much wider variety of suitable significance levels and is valid under the same conditions as the sign test. The purpose of this ~~note~~ is to investigate the power efficiency of the significance tests for the median formed in this way for the particular case in which each observation is drawn from the same normal population. For the cases considered, it is found that the power efficiency is almost always decreased by using two or more subsets rather than all the observations as a single set; sometimes this decrease in power efficiency is very large. Also, for a given significance level, it is found that the power efficiency can vary noticeably with the manner in which the test is formed.

2. Statement of conditions and tests. Consider n independent observations drawn from n (possibly different) populations which satisfy the conditions

(i) Each population has a unique median

(ii) Each population is continuous at the median (i.e. its cdf is continuous at the median).

(iii) The median of each population has the same value φ

The following significance tests compare φ with a given value φ_o through use of these n observations.

Let $x(1|n),\ldots, x(n|n)$ represent the values of the n observations arranged in increasing order of magnitude. Then the one-sided sign test of $\varphi<\varphi_o$ is defined by

$$\text{(1)} \qquad \underline{\text{Accept}}\ \varphi<\varphi_o\ \underline{\text{if}}\ x(i|n) < \varphi_o .$$

The significance level of this test is equal to

$$\text{(2)} \qquad Pr[x(i|n)<\varphi] = (\tfrac{1}{2})^n \sum_{s=i}^{n} \frac{n!}{s!\,(n-s)!} .$$

The one-sided sign test of $\varphi>\varphi_o$ is defined by

$$\text{(3)} \qquad \underline{\text{Accept}}\ \varphi>\varphi_o\ \underline{\text{if}}\ x(n+1-i|n)>\varphi_o .$$

The significance level of this test is also given by (2).

An equal tail sign test of $\varphi\neq\varphi_o$ is given by

$$\text{(4)} \qquad \underline{\text{Accept}}\ \varphi\neq\varphi_o\ \underline{\text{if either}}\ x(i|n)<\varphi_o\ \underline{\text{or}}\ x(n+1-i|n)>\varphi_o,\ (i>\tfrac{n+1}{2})$$

The significance level of this test is equal to twice the value of (2).

Now suppose that the n independent observations were drawn as r subsets, the kth subset being of size n_k, $(k=1,\ldots, r\leq n; n=n_1+\ldots+n_r)$. Then, for given α_k, $Pr[x(\alpha_k|n_k)<\varphi]$ can be computed for the kth

subset by the use of (2). Let y_j be the jth largest of $x(\alpha_1|n_1)$, $\ldots, x(\alpha_r|n_r)$. Then the values of $Pr(y_u<\varphi)$ and $Pr(y_u<\varphi<y_v)$ can be determined from elementary probability considerations by using the values of the $Pr[x(\alpha_k|n_k)<\varphi]$ and the fact that the subsets are independent. By appropriate choices of r, the n_k, the α_k, u, and v, significance tests with a wide variety of suitable significance levels can frequently be found.

The subset tests considered here are restricted to the following:

(5) $\underline{\text{Accept}}\ \varphi<\varphi_0\ \underline{\text{if}}\ \max\left[x(\alpha_k|n_k);k=1,\ldots,r\right]<\varphi_0$.

The significance level of this test equals

(6) $$\prod_{k=1}^{r} Pr\left[x(\alpha_k|n_k)<\varphi\right].$$

The one-sided test of $\varphi>\varphi_0$ is

(7) $\underline{\text{Accept}}\ \varphi>\varphi_0\ \underline{\text{if}}\ \min\left[x(n_k+1-\alpha_k|n_k);k=1,\ldots,r\right]>\varphi_0$.

The significance level of this test is also equal to (6).

The equal tail test of $\varphi\neq\varphi_0$ is defined by

(8) $\underline{\text{Accept}}\ \varphi\neq\varphi_0\ \underline{\text{if either}}\ \max\left[x(\alpha_k|n_k);k=1,\ldots,r\right]<\varphi_0$

$\underline{\text{or}}\ \min\left[x(n_k+1-\alpha_k|n_k);k=1,\ldots,r\right]>\varphi_0$.

The significance level of this test is equal to twice the value of (6).

3. Significance levels. Examination of the results of section 2 shows that the significance levels of all of the tests (1), (3), (4), (5), (7), (8) are determined if the significance levels of tests (1) and (5) are known. Thus it is sufficient to restrict significance level considerations to tests (1) and (5).

From a practical viewpoint, the important significance levels for one-sided tests are in the .05 - .005 range. Table 1 contains a list of the tests of type (1) which have significance levels near this range for $n \leq 15$. It is seen that suitable significance levels are not available for $n<4$. Also there is a very limited choice of satisfactory levels for all values of n from 4 to 15 inclusive.

If the n observations are drawn as subsets, suitable significance levels are still not available for $n<4$. For $n \geq 6$, however, a greater variety of satisfactory levels can be obtained. In practice the .05, .025, .01, .005 significance levels are of particular importance. Table 2 shows how closely these levels can be approximated for test (5). The approximations can be made very close for $n=12$ and at least as good for $n>12$.

Thus, if the observations are not drawn as subsets, the number of suitable significance levels is very limited. Drawing the observations as subsets furnishes many more satisfactory significance levels. Examination of the power efficiencies listed in Tables 1 and 2 shows, however, that noticeable efficiency can be lost by using the subset drawing procedure if the observations are a sample from a normal population.

4. <u>Power efficiency derivations</u>. The power efficiency of a significance test is defined in [1]. Essentially the power efficiency of a significance test equals 100 m/n %, where n is the sample size for the given test and m is the sample size (not necessarily integral) of the corresponding most powerful test at the same significance level whose power function is approximately the same as that of the given test.

For one-sided tests of $\rho < \rho_0$, $\rho > \rho_0$ and symmetrical tests of $\rho \neq \rho_0$, the most powerful test for a sample from a normal population (unknown variance) is the appropriate Student t-test. Also it is sufficient to limit investigations to the one-sided tests (1) and (5). As shown in [1], tests (3) and (4) have the same power efficiency as (1) while tests (7) and (8) have the same power efficiency as test (5).

The power efficiencies listed in Table 1 were obtained from [1, Table 6] and will not be derived here.

Let the normal population have variance σ^2 and consider test (5)

$$
\begin{aligned}
\text{Power Function} &= \prod_{k=1}^{r} \Pr\left[x(\alpha_k|n_k) < \rho_0\right] \\
(9) \qquad &= \prod_{k=1}^{r} \Pr\left[\frac{x(\alpha_k|n_k)-\rho}{\sigma} < \frac{\rho_0-\rho}{\sigma}\right] \\
&= \prod_{k=1}^{r} \left\{\sum_{s=\alpha_k}^{n_k} \frac{n_k!}{(n_k-s)!\,s!}\left[N(\delta)\right]^{s}\left[1-N(\delta)\right]^{n_k-s}\right\},
\end{aligned}
$$

where

$$
\delta = (\rho_0-\rho)/\sigma, \qquad N(\delta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\delta} e^{-\frac{1}{2}x^2}dx.
$$

The power function values for test (5) listed in Table 3 were computed from (9). The corresponding t-test power function values were obtained by using the normal approximation given in [2]. The power efficiencies listed in Table 2 were obtained from the results of

Table 3.

Examination of Tables 1 and 2 shows that a substantial amount of information can be lost by using a test of the form (5) rather than of the form (1). The asymptotic results of [3] show that test (1) is always at least 63.7% efficient. Some of the tests of Table 2, however, are only slightly more than 40% efficient. On the other hand, some tests of type (5) have approximately the same power efficiency as the corresponding types (1) tests. For example, the test

$$\underline{\text{Accept}}\ \rho < \rho_0\ \underline{\text{if}}\ \max\left[x(8|9), x(2|3)\right] < \rho_0$$

compares very favorably with the corresponding type (1) tests. This shows that the use of test (5) should not be completely avoided but that the power efficiency of a test of this type should be investigated before that test is considered for application.

Tests having the same significance level and based on the same total number of observations can have a wide variety of power efficiencies depending on the way in which the test is formed. Consider the case of significance level .0175 and n = 15. Table 4 contains the power efficiencies of six tests formed in different ways. These tests vary from 42% to 74% efficient. Use of the test

$$\underline{\text{Accept}}\ \rho < \rho_0\ \underline{\text{if}}\ \max\left[x(4|7), x(1|2), x(1|2), x(4|4)\right] < \rho_0$$

results in the loss of 8.7 sample values while only 3.85 sample values are lost by using the test

$$\underline{\text{Accept}}\ \rho < \rho_0\ \underline{\text{if}}\ x(12|15) < \rho_0 .$$

It is to be observed that the above power efficiency investigations are based on the assumption that the n observations are a sample from a normal population. If there is no reason to believe that the n observations come from the same normal population, the power efficiencies listed in Tables 1 and 2 may be far from the true values. The tests formed by drawing the observations as subsets might be very efficient for certain non-normal situations.

REFERENCES

[1] John E. Walsh, "Some significance tests for the median which are valid under very general conditions." To appear in Annals of Math. Stat.

[2] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution," Biometrika, Vol. 31 (1940), p. 376.

[3] John E. Walsh, "On the asymptotic power efficiency of the sign test for slippage of means." Submitted to Annals of Math. Stat.

TABLE 1

Type (1) Tests with Significance Levels Near the .05 - .005 Range

Test	Significance Level	Approx. Efficiency	Test	Significance Level	Approx. Efficiency
$x(4\|4) < \varphi_0$	.0625	95%			
$x(5\|5) < \varphi_0$	.0312	96%			
$x(6\|6) < \varphi_0$	.0156	95%			
$x(7\|7) < \varphi_0$	.0078	95%	$x(6\|7) < \varphi_0$	.0625	80%
$x(8\|8) < \varphi_0$	.0039	95%	$x(7\|8) < \varphi_0$	.0352	80%
$x(8\|9) < \varphi_0$	.0195	82%			
$x(9\|10) < \varphi_0$	.0107	80%	$x(8\|10) < \varphi_0$	.0547	75%
$x(10\|11) < \varphi_0$	.0059	81%	$x(9\|11) < \varphi_0$	.0327	76%
$x(10\|12) < \varphi_0$	.0193	75%			
$x(11\|13) < \varphi_0$	.0112	75%	$x(10\|13) < \varphi_0$	.0462	70%
$x(12\|14) < \varphi_0$	.0065	78%	$x(11\|14) < \varphi_0$	.0287	73%
$x(13\|15) < \varphi_0$	.0037	78%	$x(12\|15) < \varphi_0$	.0175	74%
$x(11\|15) < \varphi_0$	.0593	70%			

TABLE 2

Some Type (5) Tests Near the .05, .025, .01, .005 Significance Levels

n	Test Accept $\varphi<\varphi_0$ if	Significance Level	Approx. Efficiency	Test Accept $\varphi<\varphi_0$ if	Significance Level	Approx. Efficiency
4	$x(4\mid 4)<\varphi_0$	.0625	95%			
5	$x(5\mid 5)<\varphi_0$	.0312	96%			
6	$x(6\mid 6)<\varphi_0$	.0156	95%	$\max\left[x(4\mid 4), x(1\mid 2)\right]<\varphi_0$	.0469	72%
7	$x(7\mid 7)<\varphi_0$	.0078	95%	$\max\left[x(5\mid 5), x(1\mid 2)\right]<\varphi_0$	.0235	75%
	$\max\left[x(4\mid 5), x(2\mid 2)\right]<\varphi_0$	.0469	73%	$\max\left[x(5\mid 6), x(1\mid 1)\right]<\varphi_0$	.0547	75%
8	$x(8\mid 8)$	.0039	95%	$\max\left[x(1\mid 2), x(6\mid 6)\right]<\varphi_0$	.0117	77%
	$\max\left[x(4\mid 5), x(3\mid 3)\right]<\varphi_0$	.0235	73%	$\max\left[x(2\mid 3), x(1\mid 2), x(3\mid 3)\right]<\varphi_0$	.0469	58%
9	$\max\left[x(1\mid 2), x(7\mid 7)\right]<\varphi_0$	.0059	79%	$\max\left[x(3\mid 4), x(5\mid 5)\right]<\varphi_0$	.0098	75%
	$\max\left[x(2\mid 3), x(1\mid 2), x(4\mid 4)\right]<\varphi_0$	.0235	61%	$\max\left[x(2\mid 5), x(4\mid 4)\right]<\varphi_0$	.0508	47%
10	$\max\left[x(3\mid 4), x(6\mid 6)\right]<\varphi_0$	.0049	77%	$\max\left[x(8\mid 9), x(1\mid 1)\right]<\varphi_0$	.0098	80%
	$\max\left[x(2\mid 5), x(5\mid 5)\right]<\varphi_0$	.0254	51%	$\max\left[x(1\mid 3), x(1\mid 3), x(4\mid 4)\right]<\varphi_0$	.0469	43.5%
11	$\max\left[x(8\mid 9), x(2\mid 2)\right]<\varphi_0$	.0049	79%	$\max\left[x(2\mid 3), x(3\mid 4), x(4\mid 4)\right]<\varphi_0$	.0098	65%
	$\max\left[x(1\mid 3), x(1\mid 3), x(5\mid 5)\right]<\varphi_0$	.0235	48%	$\max\left[x(2\mid 5), x(2\mid 3), x(3\mid 3)\right]<\varphi_0$	.0508	43%
12	$\max\left[x(2\mid 3), x(3\mid 4), x(5\mid 5)\right]<\varphi_0$	.0049	66%	$\max\left[x(8\mid 9), x(2\mid 3)\right]<\varphi_0$	.0098	73%
	$\max\left[x(2\mid 5), x(2\mid 3), x(4\mid 4)\right]<\varphi_0$	.0254	45%	$\max\left[x(2\mid 5), x(6\mid 7)\right]<\varphi_0$	.0508	50%

TABLE 3

Power Function Values of Some Type (5) Tests

Near the .05, .025, .01, .005 Significance Levels

Significance Test	Sample Size	Approx. Efficiency	Significance Level	Approximate Values of Power Function δ=.6	δ=1.2	δ=1.8	δ=2.4
t	4.32		.0469	.234	.590	.881	.983
max[x(4$\mid$4), x(1$\mid$2)] < φ_0	6	72%	.0469	.256	.603	.860	.965
t	5.25		.0235	.164	.511	.851	.980
max[x(5$\mid$5), x(1$\mid$2)] < φ_0	7	75%	.0235	.186	.534	.829	.958
t	5.1		.0469	.282	.700	.947	.996
max[x(4$\mid$5), x(2$\mid$2)] < φ_0	7	73%	.0469	.306	.701	.919	.984
t	5.25		.0547	.326	.758	.968	
max[x(5$\mid$6), x(1$\mid$1)] < φ_0	7	75%	.0547	.346	.756	.947	
t	6.17		.0117	.113	.442	.821	.975
max[x(1$\mid$2), x(6$\mid$6)] < φ_0	8	77%	.0117	.135	.473	.800	.950
t	5.85		.0235	.194	.601	.915	.993
max[x(4$\mid$5), x(3$\mid$3)] < φ_0	8	73%	.0235	.222	.620	.886	.975
t	4.68		.0469	.258	.646	.919	.992
max[x(2$\mid$3), x(1$\mid$[illegible]), x(3$\mid$3)] < φ_0	8	58%	.0469	.288	.658	.891	.976
t	7.1		.0059	.080	.384	.796	.972
max[x(1$\mid$2), x(7$\mid$7)] < φ_0	9	79%	.0059	.098	.418	.771	.942
t	6.75		.0098	.123	.472	.846	.983
max[x(3$\mid$4), x(5$\mid$5)] < φ_0	9	75%	.0098	.140	.506	.826	.960

TABLE 3 (Continued)

Significance Test	Sample Size	Approx. Efficiency	Significance Level	Approximate Values of Power Function δ=.6	δ=1.2	δ=1.8	δ=2.4
t	5.5		.0235	.17[illegible]	.553	.881	.987
max[x([illegible]$\mid$3), x(1$\mid$2), x(4$\mid$4)] < φ_0	9	61%	.0235	.209	.582	.859	.967
t	4.2		.0508	.243	.594	.881	.932
max[x(2$\mid$5), x(4$\mid$4)] < φ_0	9	47%	.0508	.270	.610	.861	.966
t	7.7		.0049	.081	.413	.833	.983
max[x([illegible]$\mid$4), x([illegible]$\mid$6)] < φ_0	10	77%	.0049	.102	.443	.796	.952
t	8		.0098	.160	.631	.944	.998
max[[illegible], x(2$\mid$1)] < φ_0	10	80%	.0098	.178	.642	.92[illegible]	.989
t	5.1		.0254	.169	.512	.848	.978
max[[illegible], x(5$\mid$5)] < φ_0	10	51%	.0254	.196	.540	.830	.958
t	4.35		.0469	.237	.596	.884	.984
max[x([illegible]$\mid$3), x(2$\mid$3), x(4$\mid$4)] < φ_0	10	43.5%	.0469	.268	.611	.863	.967
t	8.7		.0049	.106	.527	.920	.997
max[[illegible], x([illegible]$\mid$2)] < φ_0	11	79%	.0049	.129	.568	.893	.981
t	7.15		.0098	.129	.527	.898	.993
max[x(2$\mid$3), x([illegible]$\mid$4), x(4$\mid$4)] < φ_0	11	65%	.0098	.158	.551	.855	.966
t	5.3		.0235	.167	.520	.855	.981
max[x(2$\mid$[illegible]), x(2$\mid$3), x(5$\mid$5)] < φ_0	11	48%	.0235	.195	.541	.832	.959
t	4.8		.0508	.283	.635	.936	.995
max[x(2$\mid$5), x(2$\mid$3), x(3$\mid$3)] < φ_0	11	43%	.0508	.303	.664	.891	.973

TABLE 3 (Concluded)

Significance Test	Sample Size	Approx. Efficiency	Significance Level	Approximate Values of Power Function			
				$\delta = .6$	$\delta = 1.2$	$\delta = 1.8$	$\delta = 2.4$
t	8		.0049	.089	.451	.864	.990
$\max\left[x(2\|3), x(3\|4), x(5\|5)\right] < \varphi_0$	12	66%	.0049	.115	.487	.824	.998
t	8.75		.0098	.184	.700	.974	
$\max\left[x(8\|9), x(2\|3)\right] < \varphi_0$	12	73%	.0098	.200	.699	.959	
t	5.5		.0254	.189	.575	.896	.990
$\max\left[x(2\|5), x(2\|3), x(4\|4)\right] < \varphi_0$	12	45%	.0254	.220	.588	.859	.965
t	6		.0508	.353	.811	.933	
$\max\left[x(2\|5), x(6\|7)\right] < \varphi_0$	12	50%	.0508	.376	.808	.973	

TABLE 4

Some Significance Tests at the .0175 Significance Level for $n = 15$

Significance Test	Sample Size	Approx. Efficiency	Significance Level	Approximate Values of Power Function $\delta = .6$	$\delta = 1.2$	$\delta = 1.8$	$\delta = 2.4$
t	11.15		.0175	.369	.924	.999	
$x(12\|15) < \varphi_0$	15	74%	.0175	.377	.916	.996	
t	8.1		.0175	.244	.764	.983	
$\max[x(7\|8), x(4\|7)] < \varphi_0$	15	54%	.0175	.276	.764	.967	
t	8.4		.0175	.248	.787	.988	
$\max[x(5\|5), x(4\|5), x(3\|5)] < \varphi_0$	15	56%	.0175	.291	.789	.947	
t	6.3		.0175	.167	.569	.906	.992
$\max[x(4\|7), x(1\|2), x(1\|2), x(4\|4)] < \varphi_0$	15	42%	.0175	.211	.594	.859	.971
t	6.6		.0175	.178	.606	.929	.996
$\max[x(3\|5), x(2\|3), x(1\|2), x(1\|2), x(3\|3)] < \varphi_0$	15	44%	.0175	.230	.640	.860	.976
t	7.1		.0175	.231	.659	.956	.998
$\max[x(2\|3), x(2\|3), x(2\|3), x(1\|2), x(1\|2), x(2\|2)] < \varphi_0$	15	47%	.0175	.245	.681	.913	.981